Developmental Differences in Rule Learning: A Microgenetic Analysis

Robert S. Siegler and Zhe Chen

Carnegie Mellon University

Trial-by-trial strategy assessments and a microgenetic design were used to examine 4- and 5-year-olds' learning of rules for solving balance scale problems. The design allowed us to examine simultaneously the contribution to rule learning of distal variables (qualities and knowledge with which children enter the learning situation) and proximal variables (processes that they execute during learning). Developmental differences in learning arose through two distal variables that were correlated with age—initial rule use and initial encoding-helping older children to execute several proximal processes—noticing the potential explanatory role of a key variable, formulating a more advanced rule, and generalizing and maintaining the rule. Joint consideration of distal and proximal influences seems likely to be generally useful for understanding learning and development. © 1998 Academic Press

Understanding developmental differences in learning is an inherently challenging enterprise. It requires surmounting both conceptual and methodological obstacles. The conceptual obstacles arise because developmental differences in learning are inherently complex. Analyzing them forces us to specify the initial competence of children of each age, their later competence, how they get from the one to the other, and how children of different ages differ in how they get from the one to the other. These last two challenges are especially formidable, because changes often occur quite rapidly and because they tend not to proceed by the most direct or logical path that we could imagine (Coyle & Bjorklund, 1997; Flavell, 1984; Kuhn, 1995; Miller & Seier, 1994; Schauble, 1996; Sternberg, 1984).

The methodological demands of understanding change are related to the conceptual ones. Many of the most powerful cognitive-psychological meth-
ods—for example, chronometric and error analysis methods—depend on aggregating performance over many trials. The meaningfulness of results obtained using such methods depends on people’s approaches remaining constant during the period over which results are aggregated. Thus, such methods are more useful for examining steady states of knowledge than for examining rapidly changing knowledge. This limitation, together with the conceptual complexity of understanding change, accounts for the facts that far more is known about thinking at particular ages than about how changes in thinking occur, and that even less is known about developmental differences in change processes.

Partially offsetting these difficulties, two relatively recent advances have enhanced our ability to study cognitive change. One advance involves validation of immediately-retrospective verbal reports of strategy use. In the past, verbal reports of cognitive activity have been viewed with considerable skepticism (e.g., Brainerd, 1973; Nisbett & Wilson, 1977). Recent data, though, indicate that on tasks that take between 2 and 30 s, both children and adults can accurately describe how they solved the immediately preceding problem. Converging evidence from solution time and error measures attests to the validity of these self reports of strategy use (Ericsson & Simon, 1993; Pellegrino & Goldman, 1989; Siegler, 1987; 1989). Asking for such self-reports has also been found to be non-reactive. For example, both McGilly and Siegler (1990) and Bray, Fletcher, and Huffman (in press) found that asking children immediately after each trial “How did you solve that problem” did not influence either their accuracy or their frequency of use of overt strategies. Obtaining valid and non-reactive trial-by-trial reports of strategy use makes studying cognitive change considerably more feasible.

Ability to study cognitive change also has been enhanced by the increasing use of microgenetic methods. These methods have three key characteristics: (1) Knowledge is assessed before, during, and after the period of rapid change; (2) The density of observations during this period is high relative to the rate of the change; and (3) Observed behavior is intensively analyzed, with the goal of inferring the changing representations and processes that gave rise to it. Such microgenetic methods, combined with trial-by-trial strategy assessments, can provide the fine-grained data needed to understand change, because they yield a trace of the rapidly shifting thinking that often characterizes cognitive growth.

It seems useful to distinguish between two classes of influences on learning that microgenetic studies can help us understand: distal influences and proximal influences. Distal influences on learning are knowledge, processes, and qualities that we can measure before people encounter experiences that promote learning. They include age, IQ, content knowledge relevant to the task that will be presented, and other properties that children bring with them to the learning situation. Most recent studies of children’s learning have focused on such distal influences. For example, Johnson and Mervis (1994)
found that children with high IQs acquired expertise about shorebirds more rapidly than peers with average IQs; Schneider, Korkel, and Weinert (1989) found that children who possessed considerable knowledge of soccer learned more about a newly described soccer game than did peers low in such knowledge; and Crisafulli and Brown (1986) found that older children transferred problem solving schema under a broader range of conditions than did younger ones.

As suggested by these examples, the category of "distal influences on learning" encompasses diverse characteristics. It includes relatively enduring and general qualities, such as IQ and personality, and also relatively transitory and specific qualities, such as knowledge about a particular domain. The diverse examples within the category of distal influences are united by two features: all are present before the learner enters the learning situation, and all are expected to influence the learning process. Microgenetic methods can provide the data needed to identify similarities and differences in the ways in which these distal variables exert their influences.

Proximal influences on learning are processes that occur within the learning situation. They include hypotheses that learners form to account for their new observations, strategies and skills that they acquire in the course of learning, changing frequencies of use of pre-existing strategies, and generalization of strategies to novel contexts.

Examining differences between pretest and posttest performance, as is done in most studies of learning, can tell us a considerable amount about distal influences but little if anything about proximal processes. In contrast, microgenetic studies can also yield information about proximal influences on learning. For example, a microgenetic study of preschoolers' addition (Stigler & Jenkins, 1989) indicated several characteristics of learning within the experimental situation: Discovery of new strategies occurred on easy as well as difficult problems, conscious awareness of having used a new strategy was associated with rate of generalization of the strategy, and discovery of a new strategy was frequently followed by regression to older, less adequate ones.

Microgenetic studies also have the potential to integrate distal and proximal influences on learning. In particular, they can reveal the specific ways in which the distal influences exercise their effects within the learning situation. We already know, for example, that older children often learn more than younger ones from a given experience. Microgenetic designs in which older and younger children with comparable initial knowledge are presented comparable experiences can illuminate the proximal processes that lead to these developmental differences in learning.

In the present study, we performed a microgenetic analysis of older and younger children's learning about balance scales. Studying developmental differences in learning in this context was of interest both because the balance scale task has been central in a number of recent theories of cognitive
development and because it allowed us to test a general analysis of rule learning. In the next two sections, we first describe the findings about children's understanding of balance scales and then the general analysis of rule learning.

The Balance Scale Task

Inhelder and Piaget (1958) devised the balance scale task as a means of examining formal operations reasoning. Several variants of the task have been used since then; one of the most common is shown in Fig. 1. The scale includes a fulcrum and an arm that can rotate around it. The arm can tip left or right or remain level, depending on how weights (metal disks with holes in the center) are placed on pegs that protrude from the arm. The task is to predict which side, if either, would go down if wooden blocks that hold the scale motionless were removed.

Since Inhelder and Piaget's pioneering research, the balance scale has become a kind of marker task in cognitive development. It has been used to illustrate the usefulness of a variety of theories in accounting for cognitive-developmental data: Piagetian (Karmiloff-Smith & Inhelder, 1974), neo-Piagetian (Case, 1985; Marini, 1992), cultural-contextualist (Tudge, 1992), symbolic information processing (Klahr & Siegler, 1978; Langley, 1987; Newell, 1990), connectionist (McClelland, 1995; Rajmakers, Van Koten, & Molenaar, 1996), information integration (Surber & Gzesh, 1984; Wilkening & Anderson, 1982; 1991), and psychometric (Jansen & van der Maas, 1997; Wilson, 1989). It has been used to study infants (Case, 1985), preschoolers (Siegler, 1978), school-age children (Amsel, Goodman, Savoie, & Clark, 1996; Ferretti & Butterfield, 1986), and adults (Hardiman, Pollatsek, & Well, 1986). It also has been used to examine a large range of issues, including development of means-ends analysis (Case, 1985), formation of mental models (Karmiloff-Smith & Inhelder, 1974; Zelazo & Shultz, 1989), collaborative problem solving (Damon & Phelps, 1988; Tudge, 1992), and influences on rule use (Ferretti, Butterfield, Cahn, & Kerkman, 1985; Wilkening & Anderson, 1982; 1991).

One reason for the widespread interest in balance scales is the simple, hierarchically related sequence of rules through which children of different
ages progress on the task (Siegler, 1976). Some 4-year-olds and most 5-year-olds base predictions on the relative weight on the two sides of the fulcrum (Rule I). If one side has more weight, these children predict it will go down; if the weights are equal, they predict that the scale will balance. By age 8 or 9, most children use Rule II, in which they consider distance from the fulcrum if the amount of weight on the two sides is equal, but continue to rely on weight if the amounts of weight differ. By age 12 or 13, most children use Rule III, in which they always consider both weight and distance, but do not know how to resolve conflicts in which one side has more weight and the weight on the other side is farther from the fulcrum. Neither they nor adults often rely on the most advanced approach (Rule IV), in which such conflicts are resolved by computing the torques on the two sides (multiplying the weight on each side by its distance from the fulcrum).

Because of the tendency of children to form relatively simple rules for solving balance scale problems, and the consistent developmental sequence in which the rules appear, the task has proved particularly useful for examining rule learning. One factor that has been found to influence learning of balance scale rules is the discrepancy between existing knowledge and target knowledge. Children who use Rule I are more likely to form a more advanced rule if presented problems that can be solved by Rule II than if presented problems whose solutions require Rule III (Siegler, 1976).

Age also is related to learning, above and beyond the influence of initial rule use. Thus, when 5- and 8-year-olds who initially used Rule I were presented feedback on problems whose solutions demanded Rule II or III, the 8-year-olds were more likely to form such rules (Siegler, 1976).

These age-related differences in learning are closely related to differences in encoding of the relevant dimensions. To assess encoding of balance scale problems, Siegler (1976) presented children with configurations of weights on pegs and gave them 10 or 15 seconds to study the configuration so that they could "make the same problem." Then an opaque barrier was placed between the child and the original balance scale, and the child was asked to reproduce the original configuration on an identical balance scale that did not yet have any disks on its pegs. Both 5- and 8-year-olds who had earlier used Rule I to predict which side would go down were quite successful in placing the correct number of disks on each side, thus indicating that they encoded weight. The 8-year-olds, but not the 5-year-olds, also usually placed the disks on the appropriate pegs, indicating that they encoded distance despite not using the information to predict which side would go down.

Particularly compelling evidence of the role of encoding in learning came from an experiment in which 5-year-olds were taught to encode distance as well as weight (Siegler, 1976, Experiment 3). Children who received this encoding instruction subsequently benefited from feedback on problems that did not help uninstructed peers to form more advanced rules.

Similar developmental differences in learning, and relations of these dif-
ferences in learning to differences in encoding, have been found with younger children. Feedback experience with problems with different amounts of weight on the two sides of the fulcrum allowed 4-year-olds but not 3-year-olds to form Rule I; assessments of encoding indicated that the 4-year-olds already encoded weight, whereas the 3-year-olds did not; and teaching 3-year-olds to encode weight allowed them to benefit from the feedback experience that had not benefited untrained peers (Siegler, 1978).

The data from these experiments did not allow examination of whether age and encoding generated separable effects, or whether the entire effect of age was due to its relation to encoding. The data did indicate, however, that initial rule use, encoding, and perhaps age, influence rule learning.

Since the 1970s, when these studies of children’s learning about balance scales were conducted, it has become apparent that learning is more complex than simple substitution of more advanced rules and encodings for less advanced ones. Children often use multiple strategies to solve a given problem. This variability of thinking is evident within individual children and even in the gestures and speech of a given child solving a given problem (Alibali & Goldin-Meadow, 1993; Coyle & Bjorklund, 1997; Graham & Perry, 1993). Strategic variability tends to be particularly high during learning; even after children first use an approach that is conceptually quite advanced, they frequently continue to use less advanced approaches as well (Siegler, 1995; Siegler & Jenkins, 1989). Such variability also is positively related to learning: the more pretest variability shown by both children and adults, the greater the amount of learning. This relation has been documented on number conservation problems (Church & Goldin-Meadow, 1986; Siegler, 1995), mathematical equality problems (Alibali & Goldin-Meadow, 1993; Perry, Church & Goldin-Meadow, 1988), and gears problems (Perry & Elder, 1997). The findings suggest that the process by which children learn about balance scales is more complex than that hypothesized by Siegler (1976). More generally, the findings indicate a need for a general analysis of rule learning.

Components of Rule Learning

How might children generate new rules for solving problems? Four processes that seem likely to be important components of rule learning in many situations are (1) noticing potential explanatory variables; (2) formulating more advanced rules; (3) generalizing the rules to new problems, within the same situation; and (4) maintaining the rules in different, less supportive situations. Each of these processes poses challenges for learners, and failure to successfully execute any one of them can derail the rule acquisition process. Below, we consider each of these proximal influences on learning.

The first component in the sequence, noticing potential explanatory variables, is the recognition that one or more previously-unattended dimensions may be relevant to the task. After people realize that a variable is relevant,
it often is almost impossible for them not to perceive variations in it. Even imagining how other people could ignore it can be difficult. Identifying relevant variables, however, is often easier in retrospect than in prospect. Findings from chess (Chase & Simon, 1973), memory for dinosaurs (Chi, 1978), physics misconception problems (McCloskey & Kaiser, 1984; Perry & Elder, 1997), time judgments (Siegler, 1983), and the balance scale task indicate that both children and adults often fail to encode relevant dimensions. Encoding a variable is necessary but not sufficient for noticing its potential explanatory importance. Noticing requires not only encoding information about the variable, but also recognizing that variations in it could account for observed outcomes. New variables seem especially likely to be noticed when observations conflict with expectations and people search for explanations of the observations.

A second essential step in rule learning is formulating a predictive rule that includes the noticed variable. Children might immediately incorporate a newly-noticed variable into such a rule (e.g., "When weights are equal, the side with disks farther from the fulcrum goes down"). However, if children only have a vague understanding of the variable’s influence, they might not take this second step. For example, on the balance scale task, they might notice that disks are located at different distances from the fulcrum but conclude only that "The scale tips when the disks are in different places." Thus, formulating a rule that makes specific predictions based on the values of the noticed variable is a second necessary step in rule formation.

A third critical step involves generalizing the rule to novel problems. As noted earlier, after formulating a more advanced approach, children often continue to use less advanced ones as well. Such lack of generalization occurs even when the new approach is clearly more advanced than earlier ones, as in relying on the type of transformation rather than on the length of the rows to solve number conservation problems (Siegler, 1995).

The fourth and final step is to maintain the new rule under less facilitative circumstances. Maintaining a new approach is easier on problems similar to those on which it was discovered than on superficially dissimilar ones (Goswami, 1992; 1995). Maintaining also is easier during the same session than after a delay (Siegler & Stern, in press). It also seems likely to be easier to maintain a new rule when feedback is provided concerning the new approach’s superiority than when it is not.

These four components—noticing potential explanatory variables, formulating predictive rules that incorporate the variables, generalizing the rules to new problems, and maintaining them under less supportive conditions—seem likely to describe rule learning on a wide range of problems. Tracking the path of learning as children execute, or fail to execute, each component can increase the precision and depth of our understanding of rule acquisition and of developmental differences in learning. In the present study, we examined the applicability of this analysis to the balance scale task.
The Present Study

In this study, we presented 4- and 5-year-olds with a pretest, feedback problems, and a posttest. Both the pretest and posttest included two parts: a predictions test and an encoding task. These tasks were used to assess children's rule use and their encoding of weight and distance information.

In the feedback phase, half of the children of each age received feedback experience with weight problems, which was intended to help them learn Rule I if they did not already know it. The other children received feedback on distance problems, intended to help them learn Rule II, which almost none of the children knew. On each problem during the feedback phase, children were asked to predict which, if either, side would go down. Then they saw the blocks removed and the scale rotate in the direction that had more torque. Finally, they were asked “Why do you think that side went down?” Requests for such explanations were intended to simulate, in an intensified way, a common condition of learning. People often observe unexpected events and are asked, or ask themselves, variants of the question that we asked children in the experiment, “Why did that happen?” For example, if a girl saw two people on a teeter-totter, and the side with the lighter person went down, the girl might ask herself “Why did that side go down?”

This feedback phase procedure allowed measurement of each of the four hypothesized learning components. Consider how this was done in the distance problems condition (the condition in which children received feedback on distance problems). Children were said to have noticed distance if they ever cited it in their explanation of why the observed outcome occurred. Such an explanation, together with a prediction on the next trial that the side with its disks farther from the fulcrum would go down, was taken as evidence that the child had formulated a distance rule. Children were credited with generalizing the rule if, after they met the criterion for formulating the rule, they used it on at least 80% of subsequent distance problems. Finally, they were said to maintain the new rule if they continued to use it on the posttest, where no feedback was given.

This design allowed us to test four sets of predictions. One set of predictions involves the relative effects of feedback on weight and distance problems. We predicted that a higher percentage of children of both ages who did not already use Rule I would learn the rule from feedback on weight problems than would learn Rule II from feedback on distance problems. For children who did not use any rule, Rule I was closer to their existing knowledge, a condition that generally promotes learning. In addition, rules that include only a single dimension, such as Rule I, are generally easier for 4- and 5-year-olds to learn than rules that incorporate multiple dimensions, such as Rule II (Siegel, 1996).

A related, more specific prediction was that the difference in learning in the two conditions would be greater for younger than for older children.
Halford (1993) found that when older and younger preschoolers are given relevant experience with inductive and deductive inference problems, the older preschoolers are better at incorporating multiple dimensions into predictive rules. Rule II requires use of multiple dimensions (weight and distance); Rule I does not. Thus, we expected that 5-year-olds in the present study would show some learning in the distance problems condition, as well as considerable learning in the weight problems condition, whereas the 4-year-olds’ learning would be largely or totally limited to the weight problems condition.

A second set of predictions concerns relations between pretest and posttest performance (the distal relations referred to earlier). These can be illustrated most easily in the context of learning of Rule II in the distance problems condition. We expected that three pretest variables—initial rule, encoding of distance, and age—would independently predict posttest use of Rule II.

First, with regard to pretest rule use, children who already knew Rule I were expected to learn Rule II more often than children who did not yet know Rule I. One reason for this prediction was that children who already knew Rule I knew one component of Rule II (if one side has more weight, and distances are equal, the side with more weight will go down). The other reason for the prediction was that children who used Rule I on the pretest already had the general idea that systematic rules could be used to predict the outcomes of balance scale problems.

A second prediction regarding relations between pretest and posttest performance was that pretest encoding of distance would predict posttest use of Rule II, independent of the effect of initial rule use. The reason was that children who correctly encoded the locations of the disks would be more likely to hypothesize that distances from the fulcrum were related to which side went down.

A third prediction regarding pretest–posttest links was that age would contribute to learning, above and beyond the effects of initial rule use and encoding of distance. One reason for this prediction was that older children have more cognitive resources with which to process and store data and to form hypotheses regarding multiple factors (Case, 1985; 1992). The other reason was that older children would have had more opportunity to learn that considering multiple dimensions is usually necessary to explain physical events.

A third set of predictions involved relations between pretest performance and the components of rule learning that emerged during the learning phase. We expected that the best pretest predictors of noticing would be encoding of distance and age. The reasons for this prediction were the same ones that led us to expect that these pretest variables would predict use of Rule II on the posttest. On the other hand, among children who, during the learning phase, noticed the potential role of distance, we expected pretest rule use to be the best predictor of formulating and generalizing the distance rule during the learning phase. The general knowledge that using systematic rules is
useful for solving balance scale problems seemed likely to facilitate formulating a rule in the first place and then using it consistently across problems.

The fourth set of predictions involved relations among the components of rule learning. For each learning component, the immediately preceding component was expected to be the best predictor. Thus, with regard to Rule II, noticing the potential role of distance was expected to be the best predictor of formulating a distance rule, formulating a distance rule was expected to be the best predictor of generalizing it, and generalizing reliance on distance was expected to be the best predictor of using Rule II on the posttest. Each step seems necessary, or nearly so, for its successor.

METHOD

Participants

The children were 70 4-year-olds (M = 55.4 months; range = 49–59 months) and 70 5-year-olds (M = 65.1 months; range = 60–74 months). Half of the children of each age were assigned to the weight problems condition and half to the distance problems condition. All children were recruited from four Pittsburgh-area preschools that serve predominantly lower-middle class and middle class populations. Approximately equal numbers of boys and girls at each age level and from each school were assigned to the two conditions.

Apparatus and Problems

The materials were 2 wooden balance scales, 28 ring-shaped metal disks, and 4 wooden blocks. Each balance scale's arm was 32 in. long, with five pegs on each side of the fulcrum. The first peg on each side was 3 in. from the fulcrum, and each subsequent peg was 3 in. from the preceding one. The arm of the scale could tip left or right or remain level, depending on how the metal disks were placed. Each disk weighed 1.3 ounces, measured 1 in. in diameter, and had a hole in its center so that it fit on the pegs. The arm of the scale was held steady by two wooden blocks, one placed under each end.

Children were asked to predict which side would go down, or whether the scale would stay level, if the blocks were removed. For each problem, one to four weights were on a peg on each side of the fulcrum. The weights were always placed on one of the four pegs nearest to the fulcrum on their side. One scale was used for the prediction task; two were required for the encoding task.

Children were presented six types of problems:

1. Balance: Equal number of weights on each side, placed an equal distance from the fulcrum.
2. Weight: Unequal numbers of weights on the two sides, placed an equal distance from the fulcrum.
3. Distance: Equal number of weights on the two sides, placed different distances from the fulcrum.
4. Conflict-weight: More weight on one side, more distance on the other, and the side with the greater amount of weight would go down if the blocks were removed.
5. Conflict-distance: Similar to conflict-weight problems, but the side with disks more distant from the fulcrum would go down if the blocks were removed.
6. Conflict-balance: Similar to the other conflict problems, but the scale would remain balanced if the blocks were removed.
**Procedure**

Children participated in a three-phase procedure—pretest, feedback, and posttest—over three successive days. On Day 1, children received the pretest; on Day 2, they were presented the first 12 problems of the feedback phase; on Day 3, they were given the last 4 problems of the feedback phase and the posttest.

**Pretest.** Children were brought individually to a quiet room, asked to sit across the table from the experimenter, and presented the following instructions: “Today we are going to play a game with a balance scale. The scale has these pegs that are all the same space from each other. We have these pieces of metal that all weigh the same. We also have these two blocks. If we put the blocks here (one under each end of the arm), the balance scale won’t move.” The children were encouraged to hold the weights and examine the scale.

The pretest included a predictions task and an encoding task. The predictions task was composed of three weight problems, three distance problems, and three conflict-distance problems. The experimenter introduced it with the following instructions: “I’ll put the weights on the pegs in different ways and you tell me whether this side will go down or that side will go down or whether they will stay level if I take the blocks away. The balance scale won’t actually move because the blocks are here. But you tell me how the scale would move if the blocks were not here.” For each configuration of weights on pegs, the child predicted what would happen if the blocks were removed, after which the experimenter responded, “Okay.” No feedback regarding the correctness of the prediction was provided. The child was then asked to explain his/her prediction: “Why do you think this side would go down?” Although children had unlimited time to make their predictions, the overwhelming majority of children made their predictions within a few seconds of being presented the problem.

After the predictions pretest, children were presented the eight-problem encoding pretest. The experimenter told the child,

The idea of this game is for you to look at how the weights are set on the pegs of my balance scale and then to make the same problem by putting the weights on the pegs of your balance scale. First, I’ll put the weights on the pegs of my scale. After I take this board away, you look closely to see how the weights are set on the pegs. Then I’ll put the board back up so you can’t see my scale, and you put the weights on the pegs of your scale in the same way that you saw them on mine.

An example was given before the first trial. Children were allowed 10 s per trial to observe the initial configuration of weights on pegs before the scale was placed between the two scales; children then tried to reproduce the configuration on their own scale. The 12 encoding problems on the pretest included 3 weight, 3 balance, 3 distance, 1 conflict-weight, 1 conflict-balance, and 1 conflict-distance items. Again, no feedback was given.1

**Feedback phase.** When they returned to the experimental room on the second day of the study, children were told,

I’ll put the weights on the pegs in different ways, and you tell me whether this side will go down or this side will go down or whether they will stay level if I take the blocks away. The balance scale won’t actually move now, but you tell me how the scale would move if the blocks were not here. After that, I’ll take the blocks away, and we can see whether you were right or not.

---

1 An alternative means of assessing encoding of weight and distance would have involved presenting children with photos of various scale configurations to determine whether children could recognize the correct configuration of weights on pegs. Such a task might be more sensitive in assessing young children’s encoding than the present task, which required children to reproduce the configurations.
After each prediction, children saw what happened when the blocks were removed. They also received verbal feedback ("You were right, this side went down" or "No, that wasn’t right. That side went down"). The experimenter then asked the child to provide an explanation of the outcome he/she observed: "Look at the balance scale very carefully and see whether you can figure out why this side went down."

In the weight problems condition, children were presented 10 weight problems and 6 balance problems. More weight than balance problems were used because weight problems were expected to be more useful for learning the central understanding within Rule I, the importance of the relative weight on the two sides of the fulcrum. In the distance problems condition, children were presented 12 distance problems, 2 weight problems, and 2 balance problems. The reason for presenting some feedback problems that were not of the predominant type was to prevent children from forming incorrect rules that were consistent with the outcomes they observed. For example, if they had seen only distance problems, they might have formed the rule, "Greater distance goes down, otherwise guess," rather than Rule II.

Posttest: The posttest included the same predictions and encoding tasks as the pretest. The only difference was that the posttest predictions task was longer; it included 18 problems, 3 of each type. Including all six types of problems on the posttest allowed us to assess whether children constructed rules beyond Rule II.

RESULTS

The results are reported in three main sections. In the first section, we describe pretest performance and changes from the pretest to the posttest. The goal is to provide an overview of the amount of learning that occurred during the experiment. In the second section, we examine children’s changing performance during the feedback phase. Here, the goal is to provide a precise depiction of the course of learning. In the third section, we examine how individual children’s performance during each of the three phases was related to their performance in other phases. The goal of this section is to examine the paths that led to some children learning and others not, that is, to examine individual differences in learning. All reported effects are significant beyond the .05 level, unless otherwise noted. Two raters independently coded 15 children’s performance (a total of 645 predictions, 240 explanations, and 360 encoding problems). They agreed on over 99% of trials for each of the three measures.

Changes from Pretest to Posttest

Rule use. Children’s predictions were classified into one of four categories: Rule II, Rule I, Rule I’, or No Rule. Each approach is described in Table 1.

The basic criterion for using a rule was that at least 7 of 9 pretest responses or 15 of 18 posttest responses conform to the predictions of that rule. Additional criteria also had to be met when only a few problems discriminated between two rules. In particular, being classified as using Rule II required not only meeting the overall classification criterion, but also predicting on at least 2 of 3 distance problems that the side with its weight farther from the fulcrum would go down. Classification as using Rule I required meeting the overall criterion and also predicting "balance" on at least 4 of 6 of the
problems with equal amounts of weight on the two sides (distance and balance problems). Because there were several forms of Rule I', several different specific criteria were used for its classification. One variant of Rule I' was "lighter side down"; the specific criteria for this variant were at least 80% predictions of "lighter side down" on problems with unequal weight, and at least 67% predictions that the scale would balance when weight on the two sides was equal. Another variant of Rule I' was, "equal weights balance." To be classified as using this variant of Rule I', children needed to predict "balance" on at least 80% of trials with equal amounts of weight on the two sides, and to follow no clear pattern when weights were unequal. A third form of Rule I' involved predicting that the heavier side would go down when weights were unequal, but guessing that one side or the other would go down when weights were equal. The criteria here were at least 80% predictions that the side with more weight would go down when the two sides had different amounts of weight, and at least 67% predictions that the scale would not balance when weights on the two sides were equal.

Four- and five-year olds who were classified as using the same rule (e.g., Rule I) met the criteria for rule use equally strongly. For example, among children who were categorized as using Rule I, 96% of 5-year-olds' predic-

---

**TABLE 1**

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule II</td>
<td>Greater weight → down; Equal weight, equal distance → balance; Equal weight, unequal distance → greater distance down.</td>
</tr>
<tr>
<td>Rule I</td>
<td>Greater weight → down; Equal weight → balance.</td>
</tr>
<tr>
<td>Rule I'</td>
<td>Less weight → down; Equal weight → balance; or Greater weight → down; Equal weight → guess; or Equal weight → balance, Unequal weight → guess.</td>
</tr>
<tr>
<td>No Rule</td>
<td>No apparent consistency to predictions.</td>
</tr>
</tbody>
</table>
TABLE 2
Percent of Children Using Each Rule on Pretest

<table>
<thead>
<tr>
<th>Rule</th>
<th>Weight condition</th>
<th>Distance condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds</td>
<td>5-year-olds</td>
</tr>
<tr>
<td>Rule II</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Rule I</td>
<td>28</td>
<td>54</td>
</tr>
<tr>
<td>Rule I'</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>No Rule</td>
<td>57</td>
<td>24</td>
</tr>
</tbody>
</table>

tions and 94% of 4-year-olds’ predictions conformed to the predictions of the rule. The percentages were similar on both pretest and posttest.

An analysis of pretest performance indicated that 4- and 5-year-olds’ rules differed considerably at the outset of the experiment, $\chi^2(3) = 20.74$. As shown in Table 2, most 4-year-olds could not be classified as using any rule, whereas most 5-year-olds used Rule I on the pretest.

An analysis of posttest rule use indicated that 4- and 5-year-olds who did not use Rule I or II on the pretest responded similarly to experience with weight problems. As shown on the left side of Table 3, 56% of the 4-year-olds in the weight problems condition who did not use Rule I on the pretest used it on the posttest (14 of 25), versus 69% of 5-year-olds (9 of 13).

In contrast, reactions of the older and younger children to experience with distance problems differed considerably. As shown on the right side of Table 3, only 6% of 4-year-olds (2 of 35) in the distance problems condition learned that distance influenced which side would go down when distances were unequal. In contrast, 45% of 5-year-olds (15 of 33) showed such learning. Of the 15 who learned about the role of distance, 12 (36%) used Rule II on the posttest; the other 3 consistently relied on distance when distances were unequal but did not rely consistently on weight when weights were unequal.

TABLE 3
Percent of Children Using Each Rule on Posttest

<table>
<thead>
<tr>
<th>Rule</th>
<th>Weight condition*</th>
<th>Distance condition*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4-year-olds</td>
<td>5-year-olds</td>
</tr>
<tr>
<td>Rule II</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rule I</td>
<td>56</td>
<td>69</td>
</tr>
<tr>
<td>Rule I'</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>No Rule</td>
<td>20</td>
<td>23</td>
</tr>
</tbody>
</table>

* Percentages are based on 25 4-year-olds and 13 5-year-olds who did not use Rule I or II on the pretest.

* Percentages are based on 35 4-year-olds and 33 5-year-olds who did not use Rule II on the pretest.
and distances were equal. None of the 4-year-olds consistently relied on distance to solve all types of problems. Differences between the number of 4- and 5-year-olds who used Rule II were not significant on the pretest, but they were significant on the posttest, \( \chi^2(1) = 16.3 \).

The comparable learning of Rule I by 4- and 5-year-olds in the weight problems condition indicated that 4-year-olds were not in any absolute sense less good learners than 5-year-olds, or even less good at learning about balance scales. What, then, led to the 4-year-olds’ less good learning in the distance problems condition? An additional analysis was conducted to distinguish between two interpretations. One interpretation was that 4-year-olds learned less well because their previous knowledge was farther from Rule II. Most 4-year-olds used no rule or Rule I on the pretest, whereas most 5-year-olds used Rule I; Rule II is closer to Rule I than to no rule, both in the developmental sequence and in its specific content; therefore, 5-year-olds had less to learn than 4-year-olds. The other interpretation was that Rule II was more difficult for 4-year-olds to learn, above and beyond the different distances between their existing knowledge and the target rule, because it required integrating two dimensions (weight and distance).

Two analyses suggested that 4-year-olds did encounter difficulty in learning Rule II beyond that attributable to their less advanced pretest approaches. One analysis involved comparing the learning of those 4- and 5-year-olds in the distance problems condition who used the same pretest rule (Rule I). Among these 20 5-year-olds and 11 4-year-olds, the 5-year-olds were more likely to learn Rule II (55% vs 18%, \( \chi^2(1) = 3.93 \)).

Converging evidence was provided by results of a pair of analyses that compared the likelihood of children of a given age moving forward one rule in the two experimental conditions. The 5-year-olds were equally likely to move forward one rule in the weight problems and distance problems conditions, but 4-year-olds were more likely to do so in the weight problems condition. Specifically, 62% of 5-year-olds in the weight problems condition who used no rule on the pretest used Rule I on the posttest, and 55% of those in the distance problems condition who used Rule I on the pretest used Rule II on the posttest. In contrast, 55% of 4-year-olds in the weight problems condition succeeded in moving from no rule on the pretest to Rule I on the posttest, but only 18% of peers in the distance problems condition moved from Rule I to Rule II, \( \chi^2(1) = 3.93 \). Thus, 4-year-olds encountered difficulty learning Rule II above and beyond that predictable from their pretest rule use.

**Accuracy of predictions.** Analyses of percent correct predictions yielded a picture similar to the depictions yielded by the rule assessments. To determine whether children in the two conditions performed comparably on the pretest, a 2 (Feedback Condition: weight problems or distance problems) × 2 (Age: 4- or 5-year-olds) × 3 (Problem Type: weight, distance, or conflict-distance) ANOVA, on number of correct predictions was conducted. This
analysis revealed no main effect for condition and no interactions involving the condition variable, thus indicating that children in the two conditions were comparable at the time of the pretest. There was a main effect for problem type, $F(2, 276) = 58.32$, which reflected weight problems being solved more often than the other two types of problems. There also was an interaction between age and problem type, $F(2, 276) = 14.73$, which reflected 5-year-olds predicting more accurately than 4-year-olds on weight problems, $t(138) = 2.77$, but not on the other two types of problems.

To examine pretest–posttest changes in accuracy, we next conducted a 2 (Phase: pretest or posttest) × 2 (Age) × 2 (Feedback condition) × 3 (Problem type) ANOVA on percent correct answers on the three types of problems that were presented on both pretest and posttest: weight, distance, and conflict-distance. Main effects were present for all four variables. Percent correct was higher on the posttest than on the pretest, $F(1, 136) = 7.85$; on weight problems than on distance and conflict-distance problems, $F(2, 272) = 157.57$; among 5-year-olds than 4-year-olds, $F(1, 136) = 6.98$; and among children in the distance problems condition than among those in the weight problems condition, $F(1, 136) = 9.99$.

Significant 2-way interactions were present between phase and condition, $F(1, 136) = 19.50$, phase and age, $F(1, 136) = 4.98$, problem type and condition, $F(2, 272) = 6.82$, and problem type and age, $F(2, 272) = 3.13$. In addition, two 3-way interactions were present. The condition × phase × problem type interaction, $F(2, 272) = 21.04$ reflected differing patterns of changes in the two experimental conditions. In the weight problems condition, accuracy increased from pretest to posttest on weight problems (63% vs 90% correct, $t(69) = 5.82$), but it decreased on both distance problems (22% vs 8%, $t(69) = 3.56$) and on conflict-distance problems (28% vs 6%, $t(69) = 5.30$). This pattern of change reflected children in the weight problems condition basing predictions on weight more consistently on the posttest than on the pretest, which led to more accurate answers on weight problems but less accurate answers on distance and conflict-distance problems. In contrast, in the distance problems condition, percent correct increased from pretest to posttest on distance problems (8% vs 43% correct, $t(69) = 4.16$), but remained unchanged on the other two types of problems. This reflected increased use of Rule II rather than Rule I.

The age × phase × problem type interaction, $F(2, 272) = 4.74$, reflected differences in 4- and 5-year-olds’ improvements from pretest to posttest on different types of problems. The 5-year-olds’ percent correct improved from pretest to posttest on both weight problems (74% vs 85%, $t(69) = 2.34$) and distance problems (19% vs 33%, $t(69) = 2.54$). The 4-year-olds’ percent correct showed greater improvement on weight problems (56% vs 78%, $t(69) = 3.98$), but their accuracy on distance problems remained unchanged, and their percent correct on conflict-distance problems decreased (32% vs 16%, $t(69) = 3.16$). The differences between these patterns were attributable
in large part to a greater number of 5-year-olds learning Rule II, which solved distance problems correctly, and a greater number of 4-year-olds learning Rule I, which produced consistently correct performance on weight problems but consistently incorrect performance on conflict-distance problems.

These results point to a general principle: Experiences that improve performance on one kind of problem may lead to regressions on other kinds of problems. This principle suggests a further conclusion: To determine the effects of experiences on learners' understanding, their subsequent performance on diverse types of problems in the domain must be examined.

**Explanations.** We classified each child's explanation on each trial into one of four categories. Ordered from least advanced to most advanced, they were (1) *Non-explanation*: Statement did not refer to weight or distance ("I don't know"); "It's magic"); (2) *Incorrect view of weight* ("The lighter side goes down"); (3) *Correct view of weight* ("The side with more weight goes down"); (4) *Distance explanation* ("The side with its weight farther from the middle goes down").

On the pretest, 92% of explanations fell into the nonexplanation or the correct weight category. Consistent with children's predictions, an Age × Problem Type × Condition ANOVA on the number of correct weight explanations indicated main effects for age, $F(2, 272) = 11.65$, and for problem type, $F(2, 272) = 24.45$. The 5-year-olds generated more weight explanations than the 4-year-olds (68% vs 44%), and more weight explanations were advanced on weight problems than on distance or conflict-distance problems, $t(139) = 5.84$ and 2.23. No statistical analysis was conducted for the nonexplanations, since their numbers were so close to the inverses of the numbers of weight explanations, or for the incorrect weight or distance explanations, since there were so few of them.

To examine changes in explanations between the pretest and the posttest, two Age × Condition × Problem Type × Phase ANOVAs were conducted. One was on number of correct weight explanations, the other was on number of distance explanations. The ANOVA on number of weight explanations showed main effects for all four variables, four two-way interactions, one three-way interaction, and an interaction among all four variables, $F(2, 172) = 4.75$. Unlike most four-way interactions, this one had a straightforward interpretation. In the weight problems condition, the number of weight explanations increased from pretest to posttest for both age groups for all three types of problems, $t(34) > 2.0$. In contrast, in the distance problems condition, there was only a single significant pretest-posttest change. The 5-year-olds' number of weight explanations on distance problems decreased from pretest to posttest (68% vs 39%), $t(34) = 3.15$. This interaction is what would be expected if exposure to the weight problems increased likelihood of relying on weight on all types of problems, whereas exposure to distance problems lessened reliance on weight for 5-year-olds solving distance problems.
of weight was more accurate than encoding of distance; and encoding improved from pretest to posttest.

Significant interactions were present between condition and age, $F(1, 136) = 4.35$, age, phase, and dimension, $F(1, 136) = 4.25$, and condition, phase, and dimension, $F(1, 136) = 4.72$. The age $\times$ phase $\times$ dimension interaction reflected pretest-posttest improvements in encoding occurring on different dimensions for 4-year-olds and 5-year-olds. The 4-year-olds’ encoding of weight improved from pretest to posttest (46% vs 53%, $t(69) = 2.60$), whereas their encoding of distance did not improve. The 5-year-olds showed the opposite pattern. Their encoding of weight did not improve from pretest to posttest, but their encoding of distance did (24% vs 31%, $t(69) = 3.01$).

Not surprisingly, the phase $\times$ encoded dimension $\times$ condition interaction was due to encoding of each dimension improving from pretest to posttest only when the problems presented in the feedback phase varied on that dimension. Thus, encoding of weight improved from pretest to posttest in the weight problems condition (53% vs 62%, $t(69) = 2.85$) but not in the distance problems condition. Complementarily, encoding of distance did not improve in the weight problems condition, but it did improve in the distance problems condition (19% vs 24%, $t(69) = 2.54$).

Learning in the Feedback Phase

Predictions. Separate 2 (Age) $\times$ 4 (Trial block: first, second, third, or fourth quarter of feedback problems) ANOVAs were performed for the two conditions. The analysis for the weight problems condition revealed main effects for trial block, $F(3, 204) = 16.22$, and age, $F(1, 68) = 8.53$. Predictions became more accurate over trial blocks, and older children predicted the balance scale’s actions more accurately than did younger children. The interaction was also significant, $F(3, 204) = 5.98$. The 5-year-olds advanced many more correct predictions than the 4-year-olds during the first trial block (81% vs 54%, $t(68) = 4.15$), but, as shown at the top of Fig. 2, the 4-year-olds progressively caught up, until the final trial block, there was no difference (88% vs 87% correct predictions).

The ANOVA on learning in the distance problems condition was based on the three distance problems within each trial block; the weight problem within each trial block was not included, since it would not measure learning of the importance of distance. The analysis yielded main effects for trial block, $F(3, 204) = 4.03$, and age, $F(1, 68) = 4.03$ and a marginally significant interaction $F(3, 204) = 2.30$; $p = .08$. Here, the 4- and 5-year-olds’ number of correct predictions diverged over trial blocks (top of Fig. 3). Number of correct predictions was equivalent in the first trial block of the feedback phase (41% and 42%), but 5-year-olds were considerably more accurate by the final block (70% vs 46%, $t(68) = 2.40$). Thus, 4-year-olds improved more in the weight problems condition, but 5-year-olds improved more in the distance problems condition.
The parallel ANOVA on the frequency of distance explanations also yielded main effects for all four variables. In addition, all two-way interactions were significant, as were three of the four possible three-way interactions, and the four-way interaction, $F(2, 272) = 9.16$. Again, the explanation of the four-way interaction was quite straightforward. The 4-year-olds' frequency of distance explanations did not change from pretest to posttest in either condition for any type of problem. The 5-year-olds' frequency of distance explanations did increase from pretest to posttest in the distance problems condition for distance and conflict-distance problems, $r$'s(34) > 3.26. Thus, 5-year-olds in the distance problems condition learned that distance was relevant for explaining outcomes of problems on which the disks on the two sides were different distances from the fulcrum.

**Encoding.** Children's reproductions of weights on pegs were scored for accuracy of coding of weight and distance on each trial of the encoding test. A score of 1 for the weight dimension was given when the child correctly reproduced the numbers of disks on both sides of the fulcrum; otherwise, the score was 0. Similarly, a score of 1 was given for the distance dimension when the child put the disks the correct distance from the fulcrum on both sides; otherwise, the score was 0.

Pretest performance was analyzed using an Age $\times$ Condition $\times$ Encoded Dimension (weight or distance) ANOVA. The analysis yielded main effects for encoded dimension, $F(1, 137) = 222.16$, and for age, $F(1, 137) = 15.68$. On the pretest, children encoded weight correctly on 51% of trials, versus 19% correct encoding of distance. The 5-year-olds encoded 40% of trials correctly, versus 30% for 4-year-olds. Encoding on the pretest was comparable for the two experimental groups, and none of the interactions was significant.

Changes in encoding from pretest to posttest were analyzed via an Age $\times$ Condition $\times$ Encoded Dimension $\times$ Phase (Pretest or posttest) ANOVA. This analysis revealed main effects for age, $F(1, 136) = 21.21$, encoded dimension, $F(1, 136) = 284.61$, and phase, $F(1, 136) = 14.73$. As shown in Table 4, 5-year-olds encoded more effectively than 4-year-olds; encoding
FIG. 2. Changes during feedback phase in percent correct predictions (top) and percent correct explanations (bottom): Weight problems condition.

**Explanations.** Changes in explanations during the feedback phase paralleled changes in predictions performance. An Age X Trial Block ANOVA on number of weight explanations in the weight problems condition indicated a main effect for trial block, $F(3, 204) = 9.72$, and an interaction between age and trial block, $F(3, 204) = 3.66$. As shown at the bottom of Fig. 2, in the weight problems condition, the 4-year-olds' percentage of weight explanations increased from 62% in the first trial block to 85% in the final trial block, $F(3, 102) = 5.07$. In contrast, the 5-year-olds usually gave weight explanations throughout the feedback phase and did not increase the number of such explanations over trial blocks.

A parallel Age X Trial Block ANOVA on number of distance explanations
in the distance problems condition indicated a main effect for trial block, 
$F(3, 204) = 9.29$, and an interaction between age and trial block, $F(3, 204) = 3.61$. As on the predictions measure, the form of the interaction was 
the opposite of that in the weight problems condition (bottom of Fig. 3). The 
5-year-olds increased their number of distance explanations over trial blocks 
from 26% to 53%, $F(3, 102) = 8.27$. In contrast, the 4-year-olds generated 
few distance explanations in any trial block. (The fact that percent distance 
explanations was lower than percent distance predictions was probably due 
to the latter percentage being inflated by the 33% chance likelihood of generating 
a prediction in accord with distance on trials where the child was not sure which side would go down and therefore guessed).
An interesting feature of the explanations in the distance problems condition was that weight explanations persisted throughout the session, despite such explanations having no basis in what children observed. This was especially true among the 4-year-olds. Their percentage of weight explanations did not even decline over the course of the session: 32, 26, 24, and 31% in the four trial blocks. The use of weight explanations by 5-year-olds did decrease over trial blocks (34, 30, 23, and 20%; \( F(3,102) = 2.79 \)). However, even in the final trial block, 20% of their explanations indicated that the outcome was caused by the side that went down having more weight, despite the two sides clearly having the same numbers of weights and the children having been told that all weights weighed the same and having held weights in their hand to demonstrate this fact. Clearly, for many children, the belief that only weight influenced the balance scale’s actions died hard in many children.

**Components of learning.** Obtaining both a prediction and an explanation on each trial during the feedback phase allowed considerably more refined analysis of the change process than had been possible when only predictions performance was obtained. Having both kinds of data allowed us to assess each child’s acquisition of the four components of rule learning that were described in the introduction: noticing a potential explanatory variable, formulating a rule that included that variable, generalizing the rule to other problems, and maintaining the rule under less supportive conditions. At the time of the pretest, 65% of children in the weight problems condition already met the criteria for noticing variations in weight and for formulating a weight rule. Therefore, it was not feasible to analyze components of learning in the weight condition as we did in the distance condition. For this reason, we only report the componential analysis of learning for children in the distance problems condition. Each child’s attainment of each component was assessed as follows:

**Noticing.** To be said to have noticed the potential role of distance, the child must generate at least one explanation that refers to the locations of the weights. Such explanations can take either a predictive form (e.g., “The weights are farther away on this side, so it will go down”) or a nonpredictive form (e.g., “These weights are here but these weights are here”). The predictive form implies specific answers on other problems; the non-predictive form does not.

**Formulating.** To be said to have formulated a rule that incorporated the distance variables, the child must generate at least one explanation that follows the predictive form of noticing, and must predict on the next trial that the side with its disks farther from the fulcrum will go down.

**Generalizing.** The child must meet the criteria for formulating on at least 80% of trials after first formulating the rule.

**Maintaining.** The child’s posttest predictions must meet the criteria for
either Rule II or the distance form of Rule I (whichever side has its disks farther from the fulcrum will go down).

Figure 4 illustrates the course of learning for children in the distance problems condition. As can be seen, only 40% (27 of 68) ever met the criterion for noticing the potential role of distance from the fulcrum. Of those who did, about 75% (21 of 27) formulated a rule that included distance as a variable. Of those who formulated such a rule, about 70% (15 of 21) generalized the rule to the remaining problems. Finally, of those who generalized the distance rule during the feedback phase, almost 90% (13 of 15) used it on the posttest. In contrast, of the children who did not generalize the rule during the feedback phase, only 2% (1 of 53) used the rule on the posttest.

Separate analyses of the learning of younger and older children indicated that developmental differences in learning about the role of distance started early in the learning process. As shown in Fig. 5, only 17% of 4-year-olds noticed variations in distance, versus 64% of 5-year-olds. Thereafter, as shown in Fig. 5A, about 80% of 5-year-olds met each further criterion: 81% for formulating, 76% for generalizing, and 85% for maintaining. Half or more of 4-year-olds who noticed the variations in distance also met the criteria for each subsequent component (Figure 5B), though the number of 4-year-olds who noticed distance was too small for the subsequent percentages to be meaningful. Seen from another perspective, almost all of the 4-year-olds who did not learn Rule II failed even to notice distance (88%), whereas almost half of 5-year-olds who did not learn Rule II noticed distance but
failed at a later component (43%). To summarize, noticing the variation in distance was the largest impediment to learning, especially for the younger children, but difficulty with each of the subsequent components also led to a considerable number of children not mastering Rule II.

Time course of learning. The trial-by-trial data also allowed examination of when in the learning process children met the criteria for noticing and formulating. Of those children who ever met the criteria, almost half did so in the first trial block (48% for both noticing and formulating). However,
learning continued throughout the feedback phase; quite a few children who met the criteria did so for the first time in the fourth of the four trial blocks (19% for noticing; 24% for formulating).

Two-thirds of the children who formulated the distance rule did so only after generating at least two incorrect explanations of their observations (mean number of trials before first distance explanation = 3.25). When children in the distance problems condition first noticed the discrepancy between their prediction and the balance scale’s actions, most children insisted there were more weights on the side that went down. Others did not know how to resolve the conflict and simply said “I don’t know” when asked to explain why the one side had gone down. A few cited magic as the explanation.

Even after children generated a correct distance explanation, 28% of their subsequent explanations on similar problems did not cite distance. Half of the children in the distance problems condition who generated a distance-based explanation subsequently deviated from that type of explanation at least once. Also attesting to the considerable variability that characterized the acquisition process, 37% (10/27) of the children in the distance problems condition who used at least one distance explanation during the learning phase did not use Rule II or the distance-based form of Rule I on the posttest.

Two forms of noticing. A final interesting characteristic of the data from the feedback phase concerned the distinction between predictive and non-predictive forms of noticing. The predictive form of noticing was much more closely associated with posttest use of Rule II, or the distance-based form of Rule I. Of the 27 children who met the criteria for noticing, 17 used a rule that included the distance variable on the posttest. The 17 children who showed such mastery exhibited the predictive form of noticing on 93% of trials on which they noticed the potential role of distance. In contrast, the 10 children who met the criteria for noticing but who did not show such mastery on the posttest exhibited the predictive form of noticing on only 38% of the feedback phase trials on which they noticed distance. Thus, explaining the role of distance in a way that predicted outcomes of other problems was closely associated with learning.

Distal and Proximal Predictors of Learning

In this section, we first focus on the relation between individual children’s performance on the pretest and posttest. Then, we examine the path of change as children move from their pretest rule and encoding to new component understandings in the feedback phase to rule use and encoding on the posttest. For the same reason as in the above analysis of learning components, we will limit this detailed analysis of learning to the distance problems condition.

Distal predictors of posttest performance. To identify which distal variables were related to use of Rule II on the posttest, a regression analysis was conducted. Predictor variables were age (in months), number of correct
encodings of weight, number of correct encodings of distance, and rule use. Children who used Rule I on the pretest were scored 2, children who used Rule I’ were scored 1, and children who used no rule were scored 0. The two children who used Rule II on the pretest were excluded from the analysis.

All three distal variables that were expected to influence learning of Rule II contributed independent variance to its prediction. Pretest rule use accounted for 28% of the variance in whether children used Rule II on the posttest, encoding of distance accounted for an additional 14%, and age accounted for a further 4%. Older children, children whose pretest encoding of distance was more accurate, and children who used Rule I on the pretest were all more likely to learn. Together, the three pretest variables accounted for 46% of the variance in use of Rule II on the posttest.

Similar results were obtained when the three children who learned the distance form of Rule I on the posttest were counted as having learned from the feedback experience. In this analysis, pretest rule use accounted for 31% of the variance in posttest performance, encoding of distance accounted for an additional 12%, and age for an additional 5%. Thus, the three variables accounted for 48% of the variance in posttest rule use in this analysis.

**Distal predictors of proximal processes.** The same pretest variables were used to predict whether children attained each learning component. Two pretest variables predicted noticing. Age accounted for 35% of the variance, and encoding of distance for an additional 8%, in whether children noticed variations in distance on the feedback problems.

Given the intuitive relation between encoding of distance and noticing the potential explanatory importance of distance, encoding might have been expected to be very highly predictive of noticing. This was not the case, however. The two were only moderately correlated, $r = .49$, and age was a better predictor of noticing than was encoding. In addition, the relation between noticing and use of a distance-based rule on the posttest was considerably stronger than that between encoding and learning of such a rule ($r^2 = .70$ vs .47, $t(65) = 2.88$). Thus, encoding the location of weights on pegs is predictive of noticing its possible importance in accounting for the balance scale’s motions, but the two processes are far from identical.

For two other proximal processes, formulating and generalizing, neither encoding nor age were significant predictors. Instead, pretest rule use (use of Rule I’, I, or II) was closely related to them. Within the set of children who noticed variations in distance, pretest rule use accounted for 24% of the variance in which children went on to formulate a distance-based rule. Within the set of children who formulated such a rule, pretest rule use predicted 20% of the variance in generalization of the rule. Again, this made sense, because children who used a rule on the pretest would already have had the idea that rules could be used to solve balance scale problems, a key realization for both formulating a rule and generalizing it.
TABLE 5
Influences on Posttest Rule Use of Age, Pretest Rule, and Noticing (Distance Problems Conditions)

<table>
<thead>
<tr>
<th>Age</th>
<th>Pretest rule</th>
<th>% Who noticed</th>
<th>% Who used Rule II</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td>I</td>
<td>85</td>
<td>55</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>None or I'</td>
<td>31</td>
<td>8</td>
</tr>
<tr>
<td>4-year-olds</td>
<td>I</td>
<td>27</td>
<td>18</td>
</tr>
<tr>
<td>4-year-olds</td>
<td>None or I'</td>
<td>17</td>
<td>0</td>
</tr>
</tbody>
</table>

Proximal processes as predictors of subsequent processes. The hypothesized process that led to acquisition of Rule II was that one or more pretest variables would influence noticing, which in turn would influence rule formulation, which in turn would influence rule generalization, which in turn would influence maintenance of Rule II on the posttest. Regression analyses lent support to this hypothesis. Each regression analysis included all four pretest variables and any learning phase components that occurred earlier in the hypothesized sequence of component understandings than the component being predicted.

As noted previously, age and encoding of distance on the pretest accounted for 43% of the variance in whether children noticed the role of distance during the learning phase. Whether children noticed the potential role of distance accounted for 68% of the variance in whether they formulated a rule that included the distance variable; pretest rule use accounted for an additional 3%. Whether children formulated a rule that included the distance variable accounted for 63% of the variance in whether they generalized the rule, with encoding of distance adding a further 4%. Finally, generalized use of a distance-based rule during the learning phase accounted for 85% of the variance in use of Rule II on the posttest, with formulation of a distance-based rule accounting for an additional 5% of the variance. Thus, the analysis of learning components allowed for a fine-grained understanding of how individual children came to learn, or not to learn, Rule II.

Distal and proximal processes as joint determiners of learning outcomes. Considering both distal and proximal processes allowed an understanding of learning that was more comprehensive, yet also more precise, than did considering either alone. As shown in Table 5, among the distal (pretest) predictors, age and initial rule use were multiplicatively related to learning Rule II. Of the children who were both 5-year-olds and who used Rule I on the pretest, 55% used Rule II on the posttest. Neither factor alone predicted such learning; only 18% of 4-year-olds who used Rule I on the pretest learned Rule II, and only 8% of 5-year-olds who did not use Rule I on the pretest showed such learning.
What produced this multiplicative relation? One reasonable hypothesis was that the 5-year-olds who initially used Rule I and learned Rule II were more likely than other children to encode distance. There was truth to this hypothesis, but it was far from the whole story. As the regression analysis of relations between pretest and posttest performance indicated, pretest encoding of distance did predict posttest use of Rule II, independent of age and pretest rule use. However, encoding of distance did not discriminate those 5-year-olds who used Rule I on the pretest and subsequently learned Rule II from peers who also used Rule I on the pretest but did not learn Rule II. In particular, among the 20 5-year-olds who used Rule I on the pretest, 11 learned Rule II and 9 did not. The two groups did not differ in encoding of distance; percent correct encodings of distance on the pretest was 29% for those who learned Rule II and 27% for those who did not.³

A more complete understanding of the learning process was gained from examining one of the proximal processes, noticing, in conjunction with the distal influences on learning. As shown in Table 5, of the 5-year-olds who used Rule I on the pretest, 85% went on to notice the potential role of distance during the feedback phase. In contrast, only 23% of other children did so. Noticing, in turn, was essential for learning Rule II. Of those who noticed the potential explanatory importance of distance, 58% (14 of 24) used Rule II on the posttest. Of those who did not meet the criteria for noticing, 0% (0 of 44) did.

To summarize, 5-year-olds who used Rule I on the pretest almost always went on to notice the potential explanatory role of distance during the feedback phase, and most went on from there to use Rule II on the posttest. The large majority of other children did not notice the potential explanatory role of distance during the feedback phase, and none of them went on to use Rule II on the posttest. Thus, understanding learning required consideration of both the distal influences present at the time of the pretest and the proximal processes at work during the learning phase.

DISCUSSION

The combination of trial-by-trial strategy assessments and a microgenetic design allowed us to describe the learning process more precisely than is typically possible. In this discussion, we focus on three aspects of our find-

³ Although pretest levels of distance encoding were low in absolute terms for the 5-year-olds who used Rule I, they were higher than those of 4-year-olds and of 5-year-olds who did not use Rule I. Of the 45 children in these latter groups who did not learn Rule II, pretest encoding of distance was correct on only 14% of trials. On the other hand, all 3 children in the latter groups who did learn Rule II encoded distance very well on the pretest; they produced 58% correct encodings. These levels of encoding explain how pretest encoding could generally be associated with learning of Rule II, despite it not predicting the learning of the group that most often learned, 5-year-olds who used Rule I on the pretest.
ings: distal influences on learning, proximal influences on learning, and developmental differences in learning. Since the most interesting results concerned learning in the distance problems condition, this will be the reference except where otherwise indicated.

Distal Influences on Learning

The influences of three distal variables—age, pretest rule use, and pretest encoding—were examined. Previous studies of learning about balance scales had established that all three were associated with learning, but had not established whether age and encoding exercised independent influences. The present study indicated that all three variables do independently influence learning, with the three together accounting for almost half of the variance in acquisition of Rule II.

There were clear theoretical reasons why each variable should independently influence learning of this rule. Initial rule use seemed likely to be predictive of learning for two reasons. Children who already knew Rule I at the time of the pretest had already mastered some of the specific content of Rule II. They also possessed at least implicit understanding of the idea that systematic rules could be used to predict the outcomes of balance scale problems, an insight that is not present among all 4- and 5-year-olds.

There were also straightforward reasons for encoding of distance to be related to learning of Rule II. Consider probable reactions to the distance feedback problems of two children who used Rule I on the pretest, one of whom encoded distance and one of whom did not. When Rule I yielded incorrect predictions on the feedback problems, the child who encoded distance would be more likely to hypothesize that distances from the fulcrum had something to do with the discrepancy between expectation and observation. Encoding of a variable does not guarantee that the learner will hypothesize that the encoded variable might explain an unexpected outcome, but it seems sure to make it more likely.

The present results indicated that the linkage between improved encoding and learning can arise under considerably more natural conditions than those studied previously. Earlier evidence linking encoding and learning came from studies in which children received specific instruction in how to encode balance scale configurations (Siegler, 1976, 1978). Although this instructional manipulation was relevant to the theoretical issue, it was also quite unrepresentative of children’s experiences outside the lab—no one teaches children how to reproduce the locations of disks on pegs. On the other hand, children frequently are asked questions such as those presented in the present study, “Why do you think that happened?” They also ask themselves such questions. Previous studies have indicated that requests for explanations generally facilitate learning (Bielaczyc, Pirolli, & Brown, 1995; Chi, de Leeuw, Chiu, & LaVancher, 1994; Siegler, 1995). In the present context, the questions, together with the feedback problems, produced improved encoding as
well as more advanced predictive rules. These findings strengthen the case that improved encoding is related to learning in everyday situations as well as in laboratory experiments.

Finally, the superior learning of older children, above and beyond their pretest rule use and encoding, seemed likely to stem from their having greater cognitive resources with which to process and store information about the multiple influences that might be related to the unexpected outcomes. Another reason for older children's superior learning in this condition was that they would have had more opportunity to learn that considering multiple dimensions is usually necessary to explain physical events.

An additional finding about the distal variables concerned the relative ease of learning Rule I and Rule II for those with different initial knowledge. As predicted, a higher percentage of children learned Rule I than Rule II. In particular, 60% of those who did not use Rule I on the pretest used it on the posttest following feedback on weight problems. In contrast, only 20% of those who did not use Rule II on the pretest used it on the posttest following feedback on distance problems. The moderate discrepancy hypothesis explained a large part of this difference in learning. Rule II was particularly difficult to learn for children whose existing knowledge was furthest from it. Of those who used no rule or Rule I’ on the pretest, only 3% learned Rule II from exposure to distance problems. In contrast, of those who used Rule I on the pretest, 42% learned Rule II from exposure to the same problems. It clearly is easier to learn approaches that are just beyond one's current approach than to learn approaches that are further beyond it in the usual developmental sequence.

**Proximal Influences on Learning**

We proposed that rule learning involved four separable components: noticing potential explanatory variables, formulating rules that incorporated those variables, generalizing the rules to novel problems, and maintaining the rules under less facilitative circumstances. Results of the study provided considerable evidence for the usefulness of this componential analysis.

One way in which the analysis proved useful was in enabling us to trace each child’s path of learning. Those children who learned Rule II were almost invariably those who mastered each of the four components in turn. Children who ‘fell off the wagon’ at any point were almost doomed not to learn. On the other hand, mastery of one or two components was far from a guarantee of use of Rule II on the posttest. For example, noticing the potential role of distance and formulating a predictive rule were not enough. Children also had to generalize use of the rule if they were to have a strong likelihood of using it on the posttest. The fact that substantial percentages of children who had succeeded on all previous components went on to fail at each new one indicated that all four components were needed to account for the observed pattern of learning.
The componential analysis also allowed us to specify sources of developmental differences in learning. The largest source of these differences came at the very first component, noticing the potential explanatory role of the distance variable. Most 5-year-olds noticed the potential role of distance, most 4-year-olds did not. The poor learning of Rule II by 4-year-olds was foreordained by 83% of them failing even to notice that distance might be related to the balance scale's movement. Consider what this means in operational terms: Despite seeing 12 problems on which weights were equal and the side with its weights further from the fulcrum went down, and despite being asked each of the 12 times "Why do you think that side went down," 83% of 4-year-olds never indicated any suspicion that the locations of the disks had anything to do with the outcomes. Clearly, noticing the potential role of distance is harder than it looks.

But why was noticing this role so difficult? The findings were akin to those that have led to proposals that children and adults form mental models that govern the range of variables that they will consider as potentially relevant in forming rules for solving problems (Gentner & Stevens, 1983; Holland, Holyoak, Nisbett, & Thagard, 1986). People in general, and children in particular, are frequently very resistant to giving up such mental models, even when observations conflict with them (Chinn & Brewer, 1992). As Holland et al. (1986) noted, ability to modify mental models is limited by the encoding of the environment that accompanies the model. Thus, the failure of 4-year-olds and many 5-year-olds to learn from the distance problems condition, and the persistence of weight explanations despite all the evidence to the contrary, may reflect a self-perpetuating cycle in which their belief that only weight influences distance leads them not even to encode distance, which leads to their not noticing its potential importance when experience indicates that their existing rule is ineffective.

The analysis of proximal processes also revealed that substantial variability is present during learning, even on tasks like the balance scale where most children use systematic rules before and after learning experiences. Children could have made a sudden transition from Rule I to Rule II if they had an insight of the form, "Oh, when you have the same amount of weight, you should choose the side with the weight farther from the fulcrum." Some children did make sudden transitions, but for many, learning was considerably more gradual and halting. Half of the children in the distance problems condition noticed the potential role of distance during the feedback phase but soon returned to generating less sophisticated explanations of similar observations. Almost half of the children (48%) who generated distance explanations during the feedback phase did not use Rule II on the posttest.

On the other hand, it was also true that relative to other tasks that have been studied microgenetically, the balance scale task elicited a higher percentage of sharp transitions. Of those children who noticed the potential role of distance, 44% generated predictions and explanations in accord with it
on all subsequent feedback problems and used Rule II or the distance form of Rule I on the posttest. This relatively high frequency of abrupt transitions seems attributable to two features of the task: the simplicity of the new information that children needed to learn and the large improvement in accuracy on the feedback problems that the new rule yielded. Previous microgenetic studies of arithmetic, number conservation, locomotor activity, scientific reasoning, and mathematical equality seem to fit the generalization that the simpler the new approach being constructed, and the larger the gain in accuracy that it yields, the more rapid will be its consistent adoption (Adolph, 1997; Alibali & Goldin-Meadow, 1993; Kuhn, et al., 1995; Schauble, 1990; 1996; Siegler, 1995; Siegler & Jenkins, 1989).

The componential analysis used to study learning about balance scales seems applicable to a large number of other rule learning tasks. In particular, it seems applicable to any task that meets two criteria: 1) The outcome of the task is influenced by multiple separable dimensions; and 2) Children's existing rules fail to incorporate appropriately at least one of them. Many of the classic Piagetian tasks meet these criteria, among them problems involving pendulums, inclined planes, water displacement, class inclusion, time, speed, and distance (Inhelder & Piaget, 1958; Inhelder & Piaget, 1964).

The projection of shadows problem can be used to illustrate the applicability of the componential analysis to other tasks. On it, children are asked which object would cast a larger shadow on a screen if an intense light were turned on. The objects vary in size and in distance from the light source and the screen. Children between ages 5- and 7-years typically predict that the larger object will cast the larger shadow, regardless of other factors. To move beyond such thinking, children need to notice the distance of the object from the light source and/or the screen, formulate a rule based on these distances as well as on the size of the object, generalize the rule across shadows projection problems, and maintain the rule under more challenging circumstances. As this example suggests, the cycle could operate more than once. Children might first notice that objects closer to the light source tended to cast larger shadows, go through the other steps in rule learning, then notice that distance from the screen also influenced the size of the shadow, and eventually realize that the key variable was the ratio between the two distances.

As another illustration, consider the water displacement task. Here, children need to predict which of two objects would raise the level of water within a container to a greater extent. Among the variables that they might notice as potential explanations of observations of the relative displacements are the objects' volumes, weights, heights, shapes, material, and whether they sink or float. For each of these variables, children might go through a cycle of noticing, formulating, generalizing, and maintaining. Since the correct rule is disjunctive (the relevant variables are different when the object sinks from when it floats), it seems likely that a large number of such cycles...
might be observed in a single learning session. Thus, the componential analysis seems likely to be useful for studying rule learning in many situations.

**Developmental Differences in Learning**

The two experimental conditions in this study produced quite different developmental patterns. In the weight problems condition, younger and older children with comparable initial knowledge were equally good learners. The lack of difference in older and younger children's learning was not attributable to ceiling or floor effects, since slightly more than half of the children of each age learned from experience with the weight problems. This finding argues against the view that older children are invariably more proficient learners on balance scale tasks.

In contrast, in the distance problems condition, older children learned much more than younger ones. Differences in existing knowledge explained only part of the difference in learning. Among children who used Rule I on the pretest, three times as many 5-year-olds as 4-year-olds went on to use Rule II on the posttest. The question was how to explain this developmental difference in learning.

The microgenetic design used in this study allowed us to bring together distal and proximal influences on learning in a way that yielded a more complete understanding of the learning process than examining either alone. Among children who used Rule I, there was a very large difference between 4- and 5-year-olds' noticing of the potential explanatory role of distance. Only 23% of 4-year-olds who used Rule I noticed this potential explanatory role of distance, versus 85% of 5-year-olds. Most of those of both ages who used Rule I on the pretest and noticed distance during the learning phase went on to learn Rule II, 67% of 4-year-olds and 65% of 5-year-olds. However, so few 4-year-olds noticed distance that few learned Rule II. (No one who failed to notice distance during the feedback phase went on to use Rule II on the posttest.)

Examining proximal as well as distal influences on learning also clarified how each of the distal variables exercises its effects. Differences between older and younger children were most apparent in likelihood of noticing. Both 4- and 5-year-olds who noticed the potential role of distance were likely to learn Rule II, but far more of the older children noticed distance, and thus were in a position to form the rule. Encoding also exercised its effects primarily through its relation to noticing. Given the intuitive relation between encoding the positions of the disks and noticing that these positions could influence the balance scale's motions, this relation also makes good sense. The effects of initial rule use, on the other hand, seem to have been realized primarily through its influencing a different pair of proximal processes, formulating and generalizing a new rule. The value of following systematic rules to solve math and science problems seems obvious to educated adults,
but it may be far from obvious to preschoolers. Those children who followed
a systematic rule on the pretest seem likely to have appreciated the value of
such systematicity to a higher degree than peers who did not follow such a
systematic approach, and thus more likely to form and follow new, more
advanced rules.

This emphasis on identifying the proximal variables through which distal
variables exercise their effects is based on the same logic as Sternberg’s (e.g.,
1985) approach to studying intelligence. Sternberg’s goal was to identify the
sources of high and low IQ children’s differing performance on a variety of
inductive and deductive reasoning tasks. He traced the differing performance
to differential execution of both specific proximal processes, such as mapping
and application, and families of proximal processes, such as learning
components and metacomponents. For example, Sternberg and his col-
leagues found that high-IQ children especially excel at the execution of learning
components (Davidson & Sternberg, 1984; Marr & Sternberg, 1986).
The present approach of identifying the proximal variables through which
distal variables exercise their effects can be viewed as a generalized version
of Sternberg’s approach. It is not just IQ that can be examined with this
approach; any distal variable can. Effects associated with such distal vari-
ables as age, sex, cultural background, prior knowledge, and personality type
all could be examined in this way.

In summary, the developmental differences in learning that emerged in
the present study seemed to reflect both distal and proximal influences. When
5-year-olds were presented experiences that could engender learning of Rule
II, they were quite likely to notice the potential importance of distance from
the fulcrum, to formulate a rule that included that variable, to generalize the
rule to novel problems, and to maintain the rule on a posttest that was given
without feedback. In contrast, 4-year-olds were unlikely to notice the potential
role of distance, which precluded their learning of Rule II. Two processes
that were correlated with both age and learning of Rule II—encoding and
initial rule use—exercised their influence on different components of the
learning process. Encoding primarily influenced noticing; initial rule use pri-
marily influenced formulating and generalizing a more advanced rule. The
same component processes needed to learn this balance scale rule seem likely
to be essential to rule learning on many problems. Thus, considering both
distal characteristics that are present before children enter the learning situa-
tion and proximal processes that arise during learning seems essential for a
comprehensive understanding of cognitive growth.

REFERENCES
Society for Research in Child Development, 62 (3, Serial No. 251).


CHILDREN’S RULE LEARNING


Accepted May 4, 1998